

## Chapter 5 다변량 확률분포함수

변수가 두 개 이상인 경우 이를 이변량, 다변량이라 한다.

○은행 고객이 기다리는 시간( $X$ )과 서비스 받는 시간 ( $Y$ )

○바이러스 감염된 환자 수( $X$ )와 24시간 후 치료된 환자 수 ( $Y$ )

○학생들의 키( $X$ )와 몸무게( $Y$ )

○전광판에 사용된  $n$ 개 전구의 수명( $X_1, X_2, \dots, X_n$ )

이산형과 연속형 변수가 함께 있는 이변량, 다변량 분포도 가능하나 매우 복잡하고 유용성이 낮으므로 모두 연속형이거나 이산형인 경우에 한정하여 여기서는 다루기로 한다.

확률표본으로부터 계산되는 통계량(statistic, 예, 평균, 분산)은  $(X_1, X_2, \dots, X_n)$ 의 함수이며 모집단의 모수를 추정(inference, 추정과 검정)하는데 사용된다. 그러므로 통계량의 확률밀도함수는 다변량 변수의 확률밀도함수이다. 우선 설명과 이해의 편의를 위하여 이변량 변수인 경우 확률밀도함수를 먼저 설명하기로 하자..

### 5.1 결합확률밀도함수(Joint PDF)

주사위를 동시에 두 개 던지는 실험에서 한 주사위의 눈금을 확률변수  $X_1$ , 다른 주사위의 눈금을  $X_2$  라 정의하자. 이 확률실험에서 발생할 수 있는 경우의 수는 36가지이고 각각이 일어날 가능성(확률)이 동일하므로(equally likely) 확률변수  $(X_1, X_2)$ 의 결합밀도함수는 다음과 같이 정의할 수 있다.

$$p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2) = 1/36, \quad x_1 = 1, 2, \dots, 6; \quad x_2 = 1, 2, \dots, 6$$

**정의 (joint probability distribution function) 결합확률밀도함수**

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2) \text{ (이변량)}$$

$$f(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \text{ (다변량)}$$

확률밀도함수이므로 다음이 성립한다. (다변량인 경우)

$$\textcircled{1} p(x_1, x_2, \dots, x_n) \geq 0 \text{ (이산형)} \quad f(x_1, x_2, \dots, x_n) \geq 0 \text{ (연속형)} \text{ for all } x_1, x_2, \dots, x_n$$

$$\textcircled{2} \sum_{x_1} \dots \sum_{x_n} P(x_1, x_2, \dots, x_n) = 1 \text{ (이산형)}, \quad \int \int \dots \int f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1 \text{ (연속형)}$$

**EXAMPLE 5.1**

은행에 창구가 3개 있다. 2명의 고객이 서로 다른 시간에 왔고 서로 독립(모르는 사람)이라고 하자. 확률변수  $X_1$ 을 창구1에 온 고객 수,  $X_2$ 을 창구 2에 온 고객 수라 할 때  $(X_1, X_2)$  결합밀도함수를 구하시오.

| $x_2 \backslash x_1$ | 0   | 1   | 2   |
|----------------------|-----|-----|-----|
| 0                    | 1/9 | 2/9 | 1/9 |
| 1                    | 2/9 | 2/9 | 0   |
| 2                    | 1/9 | 0   | 0   |

**정의 (joint distribution function) 결합분포함수**

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(x_1, x_2) dx_2 dx_1$$

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

사건의 확률에서  $P(A \cap B) = P(AB)$  과 동일하다.

**EXAMPLE 5.2**

Example 5.1에서  $F(-1, 2)$ ,  $F(1.5, 2)$ ,  $F(5, 7)$ 을 구하시오.

0 / 8/9 / 1

**정리(THEOREM) 이변량 결합분포함수의 성질**

①  $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$

②  $F(\infty, \infty) = 1$

③ 비감소 함수(non-decreasing function)이다.

위의 정리의 증명은 obvious하다.

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c) \quad (\text{그래프로 보자})$$

**EXAMPLE 5.3**

확률변수  $(X_1, X_2)$ 의 결합확률밀도함수(joint PDF)는  $f(x_1, x_2) = 1, 0 \leq x_1, x_2 \leq 1$ .

(1)  $F(0.2, 0.4)$ 을 구하시오.

(2)  $P(0.1 \leq X_1 \leq 0.3, 0 \leq X_2 \leq 5)$ 을 계산하시오.

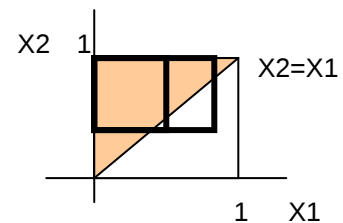
0.08 / 0.1

**EXAMPLE 5.4**

확률변수  $(X_1, X_2)$ 의 결합확률밀도함수(joint PDF)는  $f(x_1, x_2) = k(1 - x_2), 0 \leq x_1 \leq x_2 \leq 1$ .

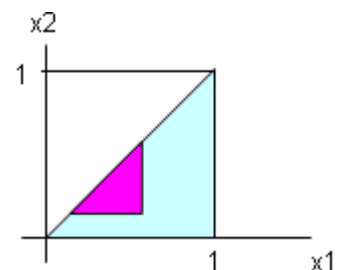
(1) 상수  $k$  구하시오.

(2)  $P(X_1 \leq 3/4, X_2 \geq 1/2)$ 을 계산하시오.

 $k = 6, 31/64$ **EXAMPLE 5.5**

확률변수  $X_1$ 은 주유소 저장 탱크에 있는 GAS 양(%)이고  $X_2$ 는 하루에 팔린 GAS 양(%)이다. 결합밀도함수는  $f(x_1, x_2) = 3x_1, 0 \leq x_2 \leq x_1 \leq 1$ 라 할 때 기름 탱크에 저장된 GAS 양이 1/2보다 작고 하루 팔린 양이 1/4 이하일 확률을 계산하시오.

5/128



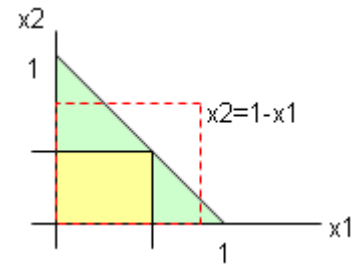
**EXAMPLE 5.6**

확률변수  $(X_1, X_2)$ 의 결합밀도함수가  $f(x_1, x_2) = 2, 0 \leq x_1 \leq 1; 0 \leq x_2 \leq 1; 0 \leq x_1 + x_2 \leq 1$ 이다.

(1)  $P(X_1 \leq 1/2, X_2 \leq 1/2)$ 을 계산하시오.

(2)  $P(X_1 \leq 3/4, X_2 \leq 3/4)$ 을 계산하시오.

1/2, 10/16

**HOMEWORK #12-1**

There are 9 executives in a company 4 are married, 3 are single, and 2 are divorced. Suppose that 3 executives are randomly selected. Let  $X_1$  denote the number of married exec. and  $X_2$  the number of never married. Find the joint pdf of  $(X_1, X_2)$

**HOMEWORK #12-2**

Let  $X_1$  denote the total time at a bank between arrival and departure and  $X_2$  the time a customer waits in line before reaching the service desk.  $f(x_1, x_2) = e^{-x_1}, 0 \leq x_2 \leq x_1$

(1) Find  $P(X_1 \leq 2, X_2 > 1)$ .

(2) Find  $P(X_1 \geq 2X_2)$ .

(3) Find  $P(X_1 - X_2 \geq 1)$  (the spent time at the service window)

**HOMEWORK #12-3**

Let  $Y_1$  and  $Y_2$  denote the proportion of time during which employee I and II perform their assigned tasks, respectively. The joint pdf of  $(Y_1, Y_2)$  is  $f(y_1, y_2) = y_1 + y_2, 0 \leq y_1 \leq 1; 0 \leq y_2 \leq 1$

(1) Find  $P(Y_1 < 1/2, Y_2 > 1/4)$ .

(2) Find  $P(Y_1 + Y_2 \leq 1)$ .

## 5.2 주변확률밀도함수 & 조건부 확률밀도함수 (Marginal and conditional pdf)

주변(Marginal)의 의미는 다른 확률변수의 모든 값에 대해 적분한다는 것이다.

정의

주변확률밀도함수:  $P(x_1) = \sum_{x_2} P(x_1, x_2)$  (이산형)  $f(x_1) = \int f(x_1, x_2) dx_2$  (연속형)

조건부확률밀도함수:  $P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)}$ ,  $f(x_1 | x_2) = \frac{f(x_1, x_2)}{f(x_2)}$  (참고:  $P(A|B) = \frac{P(A, B)}{P(B)}$ )

다변량 확률분포에도 동일한 정의가 성립된다.

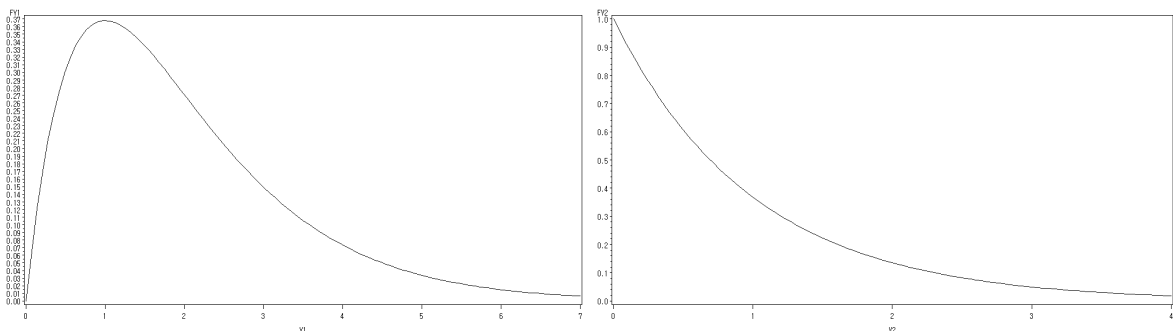
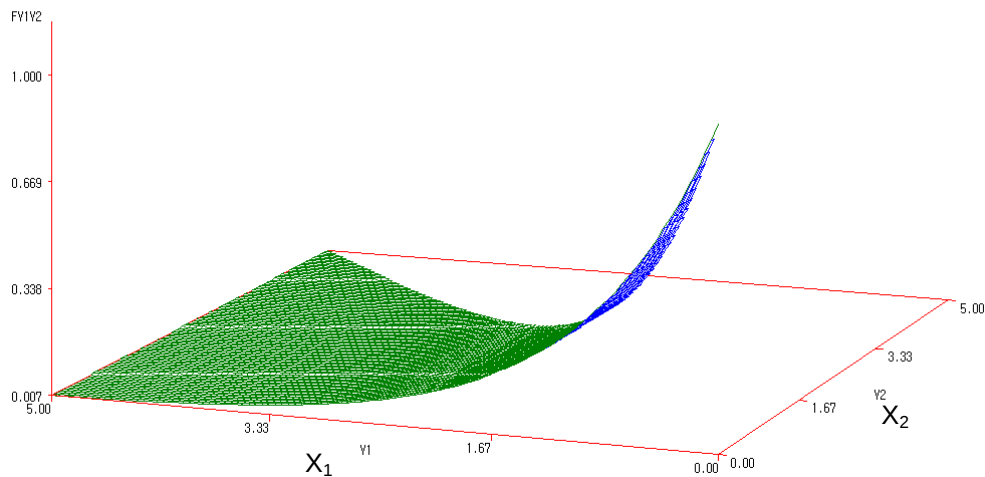
$P(x_1) = \sum_{x_2} \dots \sum_{x_n} P(x_1, x_2, \dots, x_n)$  (이산형)

$f(x_1) = \int \dots \int f(x_1, x_2, \dots, x_n) dx_2 dx_3 \dots dx_n$  (연속형) 기호:  $f_{x_1}(x_1) = f_1(x_1) = f(x_1)$



### EXAMPLE 5.7

결합밀도함수  $f(x_1, x_2) = e^{-x_1}, 0 \leq x_2 \leq x_1 < \infty$  와 그의 주변 pdf를 그리시오.



**EXAMPLE 5.8**

Example 5.1에서 (이산형)

①  $X_2$ 의 주변확률밀도함수? ② 조건부 확률밀도함수?  $P(x_1 | X_2 = 1)$

| $x_2 \backslash x_1$ | 0   | 1   | 2   |
|----------------------|-----|-----|-----|
| 0                    | 1/9 | 2/9 | 1/9 |
| 1                    | 2/9 | 2/9 | 0   |
| 2                    | 1/9 | 0   | 0   |

**EXAMPLE 5.9**

Example 5.4,  $f(x_1, x_2) = 3x_1, 0 \leq x_2 \leq x_1 \leq 1$  (연속형)

①  $X_1$ 과  $X_2$ 의 주변확률밀도함수를 구하시오.

②  $X_2$ 가 주어졌을 때  $X_1$ 의 조건부확률밀도함수  $f(x_1 | x_2)$  구하시오.

**조건부 분포함수 (conditional density function)**

$X_2 = x_2$ 가 주어졌을 때 확률변수  $X_1$ 의 분포함수는  $F(x_1 | x_2) = P(X_1 \leq x_1 | X_2 = x_2)$ 이다. 조건부 분포함수는 우선 조건부 확률밀도함수  $f(x_1 | x_2)$ 을 구한 후 분포함수를 구한다.

**EXAMPLE 5.10**

결합확률밀도함수가  $f(x_1, x_2) = 1/2, 0 \leq x_1 \leq x_2; 0 \leq x_2 \leq 2$  일 때  $P(X_1 \leq 1/2 | X_2 = 2)$ ?

$$f(x_2) = x_2 / 2, 1/4$$

**EXAMPLE 5.11**

결합 확률 밀도 함수가  $f(x_1, x_2) = 2, 0 \leq x_1 + x_2 \leq 1; 0 \leq x_1, x_2 \leq 1$  일 때  $P(X_1 \geq 1/2 | X_2 \leq 1/4)$  ?

**HOMEWORK #12-4**

The joint pdf of  $(X_1, X_2)$  is  $f(x_1, x_2) = x_1 + x_2, 0 \leq x_1 \leq 1; 0 \leq x_2 \leq 1$ .

- (1) Find the marginal pdf of  $X_1$  and  $X_2$ .
- (2) Find  $P(X_1 \geq 1/2 | X_2 \geq 1/2)$ .
- (3) Find  $P(X_1 \geq 3/4 | X_2 = 1/2)$

**HOMEWORK #12-5**

The joint pdf of  $X_1, X_2$  is  $f(x_1, x_2) = 6(1 - x_2), 0 \leq x_1 \leq x_2 \leq 1$ .

- (1) Find the marginal pdf of  $X_1$ .
- (2) Find the conditional pdf,  $f(x_2 | x_1)$ .
- (3) Find  $P(X_2 \leq 1/2 | X_1 \leq 3/4)$ .
- (4) Find  $P(X_2 > 3/4 | X_1 = 1/2)$ .

**HOMEWORK #12-6**

The joint pdf of  $X_1, X_2$  is  $p(x_1, x_2) = \frac{\binom{4}{x_1} \binom{3}{x_2} \binom{2}{3-x_1-x_2}}{\binom{9}{3}}, 0 \leq x_1, x_2 \leq 3, 1 \leq x_1 + x_2 \leq 3$ .

- (1) Find the marginal pdf of  $X_1$ . Is it a popular distribution?
- (2) Find  $P(X_2 = 1 | X_1 = 2)$ .

### 5.3 Independence (독립)

확률 이론에서 두 사건이 독립이라면  $P(AB) = P(A)P(B)$ .

#### 정의 (independence)

만약  $f(x_1, x_2) = f_1(x_1)f_2(x_2)$  (혹은  $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ ) 이라면 확률변수  $X_1$  와  $X_2$  는 서로 독립이다. 확률변수  $X_1$  와  $X_2$  는 서로 독립이라면  $f(x_1, x_2) = f_1(x_1)f_2(x_2)$  이 성립한다.  $X_1, X_2$  서로 독립  $\Leftrightarrow$  (iif)  $f(x_1, x_2) = f_1(x_1)f_2(x_2)$

#### 확률표본 (Random Sample)

$X_1, X_2, \dots, X_n$  확률표본  $\Leftrightarrow$  (iif)  $X_i \stackrel{\sim}{\text{iid}} f(x)$

IID: Independently(독립) and Identically distributed(동일한 분포)

$$f(x_1, x_1, \dots, x_n) = (\text{independently}) f(x_1)f(x_2)\dots f(x_n) = (\text{identically}) [f(x)]^n$$



#### EXAMPLE 5.12

Example 5.1, 확률변수  $X_1, X_2$  는 서로 독립인가?

No



#### EXAMPLE 5.13

결합밀도함수  $f(x_1, x_2) = 6x_1x_2, 0 \leq x_1, x_2 \leq 1$  일 경우 확률변수  $X_1, X_2$  는 서로 독립인가?

Yes



#### EXAMPLE 5.14

결합밀도함수  $f(x_1, x_2) = 2, 0 \leq x_2 \leq x_1 \leq 1$  일 경우 확률변수  $X_1, X_2$  는 서로 독립인가?

No

**정리(Theorem)**

확률변수  $X_1, X_2$  의 범위(range)가 서로 종속되지 않고 결합확률밀도함수를 두 변수의 함수의 곱으로, 즉  $f(x_1, x_2) = g(x_1)h(x_2)$ , 표시할 수 있다면 두 확률밀도함수는 서로 독립이다.. 그 역도 성립한다.

**EXAMPLE 5.15**

결합확률밀도함수  $f(y_1, y_2) = 2y_1, 0 \leq y_1, y_2 \leq 1$  일 경우 확률변수  $X_1, X_2$  는 서로 독립인가?

Yes

**HOMEWORK #13-1**

The length of life  $X$  for a bulb is distributed as  $f(x) = \frac{1}{3}e^{-x/3}, 0 < x$ . Randomly selected two bulbs have independently length of  $X_1, X_2$ . Find  $P(X_1 + X_2 \leq 1)$ .

**HOMEWORK #13-2**

Two telephone calls come into a switchboard at random times in a fixed 1-hour period. Assume that the calls are made independently of one another.

- (1) What is the probability that both calls are made in the first half hour?
- (2) What is the probability that the calls are made within 5 minutes of each other?

## 5.4 기대값(Expected value)

정의

$$E(g(X_1, X_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p(x_1, x_2) \text{ (이산형)}$$

$$E(g(X_1, X_2)) = \int \int g(x_1, x_2) f(x_1, x_2) dx_2 dx_1 \text{ (종속형)} \quad (\text{이변량})$$

$$E(g(X_1, X_2, \dots, X_n)) = \sum_{x_1} \dots \sum_{x_n} g(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n) \text{ (이산형)}$$

$$E(g(X_1, X_2, \dots, X_n)) = \int \dots \int g(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n) dx_n dx_{n-1} \dots dx_1 \text{ (연속형)} \quad (\text{다변량})$$

다변량 확률밀도함수에서  $E(X_k)$  을 구하려면 우선  $X_k$  에 대한 주변확률밀도함수를 구하고 기대값을 구하면 된다.



### EXAMPLE 5.16

결합밀도함수  $f(x_1, x_2) = 2x_1, 0 \leq x_1, x_2 \leq 1$  일 경우  $E(X_1 X_2)$ ,  $E(X_1)$ ,  $V(X_1)$  을 구하시오.

1/3, 2/3, 1/18



### EXAMPLE 5.17

결합밀도함수  $f(x_1, x_2) = 2(1 - x_1), 0 \leq x_1, x_2 \leq 1$  일 경우  $E(X_1 X_2)$  을 구하시오.

1/6

**정리(THEOREM)**

(1) 상수  $c$  에 대해  $E(cg(X_1, X_2, \dots, X_n)) = cE(g(X_1, X_2, \dots, X_n))$ .

(2)  $E(g_1(X_1, X_2, \dots, X_n) + g_2(X_1, X_2, \dots, X_n) + \dots + g_k(X_1, X_2, \dots, X_n))$   
 $= E(g_1(X_1, X_2, \dots, X_n)) + E(g_2(X_1, X_2, \dots, X_n)) + \dots + E(g_k(X_1, X_2, \dots, X_n))$

$$(예) E(X_1X_2 + X_1^2) = E(X_1X_2) + E(X_1^2)$$

(3) 만약  $X_1, X_2$  가 독립이라면  $E(X_1X_2) = E(X_1)E(X_2)$  이 성립한다.

$$E(g(X_1)h(X_2)) = E(g(X_1))E(h(X_2)) \quad (\text{함수의 곱의 경우에도 성립한다})$$

**EXAMPLE 5.18**

결합확률밀도함수  $f(x_1, x_2) = 3x_1, 0 \leq x_2 \leq x_1 \leq 1$  일 경우  $E(X_1 - X_2)$  을 구하시오.

3/4, 3/8, 3/8

**EXAMPLE 5.19**

결합확률밀도함수  $f(x_1, x_2) = 6(1 - x_2), 0 \leq x_1 \leq x_2 \leq 1$  일 경우  $E(X_1), E(X_2), E(X_1 - 3X_2)$  구하시오

1/4, 1/2, -5/4

$$f_1(x_1) = 3 - 6x_1 + 3x_1^2, 0 \leq x_1 \leq 1$$

$$f_2(x_2) = 6x_2(1 - x_2), 0 \leq x_2 \leq 1$$

**HOMEWORK #13-3**

Let  $f(y_1, y_2) = y_1 + y_2, 0 \leq y_1, y_2 \leq 1$ . Find  $E(30Y_1 + 25Y_2)$ .

**HOMEWORK #13-4**

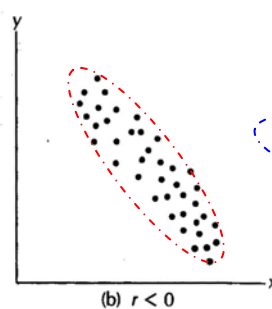
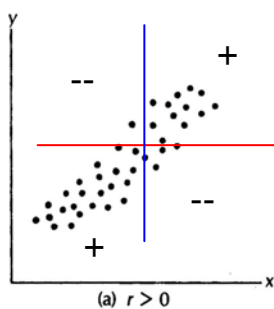
Let  $f(x_1, x_2) = 1/x_1, 0 \leq x_2 \leq x_1 \leq 1$ . Find  $E(X_1 - X_2)$ . TIP  $\int \frac{1}{x} dx = \ln x$

**5.5 공분산(Covariance)과 상관계수(correlation)**

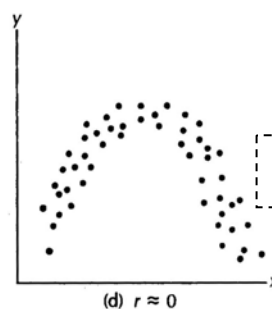
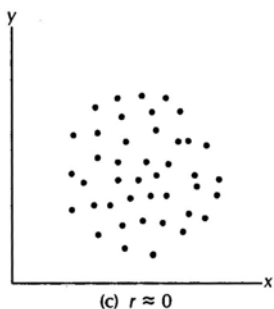
두 변수간의 직선(선형) 관계 정도를 측정하는 기대값을 공분산이라 한다. 공분산은 확률변수 값의 크기(단위)에 의존하므로 변수의 표준화와 동일한 개념으로 표준편차로 나누어준 값을 상관계수라 한다. 상관계수는 -1과 1 사이의 값이다. 즉  $-1 \leq \rho \leq 1$  이다. 표본 데이터의 상관계수는  $r$ 로 표현한다.  $\rho = 0$ 이면 상관 관계가 전혀 존재하지 않는다.  $\rho = 1$ 은 완벽한 양의 상관관계(한 변수의 값이 커지면 다른 변수 값이 커진다),  $\rho = -1$ 은 완벽한 음의 상관관계가 있다고 한다.

$$COV(X_1, X_2) = E(X_1 - E(X_1))(X_2 - E(X_2)) = E(X_1 - \mu_1)(X_2 - \mu_2) \text{ (공분산)}$$

$$\rho = \frac{COV(X_1, X_2)}{\sqrt{V(X_1)}\sqrt{V(X_2)}} = \frac{COV(X_1, X_2)}{\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}} = \frac{COV(X_1, X_2)}{\sigma_1\sigma_2} \text{ (상관계수)}$$



동일한 상관계수?



상관계수는 직선 관계 정도 측정

**정리(THEOREM)**

$$\text{COV}(X_1, X_2) = E(X_1 - E(X_1))(X_2 - E(X_2)) = E(X_1 X_2) - E(X_1)E(X_2)$$

**Proof** 기대값은 확률변수가 아니라 상수임을 이용한다.

$$\begin{aligned} E(X_1 - E(X_1))(X_2 - E(X_2)) &= E[X_1 X_2 - X_1 E(X_2) - E(X_1) X_2 + E(X_1) E(X_2)] \\ &= E(X_1 X_2) - E(X_1) E(X_2) (\because E(X_1 E(X_2)) = E(X_2) E(X_1)) \end{aligned}$$

**EXAMPLE 5.20**

결합밀도함수  $f(x_1, x_2) = 3x_1, 0 \leq x_2 \leq x_1 \leq 1$  일 경우 공분산  $\text{COV}(X_1 X_2)$  을 구하시오.

3/4, 3/8, 3/10, **0.02**

**EXAMPLE 5.21**

결합밀도함수  $f(x_1, x_2) = 2x_1, 0 \leq x_1, x_2 \leq 1$  일 경우 공분산  $\text{COV}(X_1 X_2)$  을 구하시오.

0

**정리(THEOREM)**

만약  $X_1 X_2$  가 서로 독립이면 공분산이 0이다.  $X_1, X_2$  서로 독립  $\Rightarrow \text{COV}(X_1, X_2) = 0$

위 정리의 역은 성립하지 않는다. 즉 공분산이 0이라도 독립이 아닐 수 있다.

**EXAMPLE 5.22**

| $X_2 \backslash X_1$ | -1   | 0    | 1    |
|----------------------|------|------|------|
| -1                   | 1/16 | 3/16 | 1/16 |
| 0                    | 3/16 | 0    | 3/16 |
| 1                    | 1/16 | 3/16 | 1/16 |

**HOMEWORK #14-1**

Let  $f(x_1, x_2) = 6(1 - x_1), 0 \leq x_1 \leq x_2 \leq 1$ . Find  $COV(Y_1, Y_2)$ .

**HOMEWORK #14-2**

Let the discrete random variable.  $Y_1, Y_2$  have the joint pdf  $p(y_1, y_2) = 1/3$  for  $(y_1, y_2) = (-1, 0), (0, 1), (1, 0)$ .

Find  $COV(y_1, y_2)$ . Are  $Y_1, Y_2$  independent?

**5.7 선형 함수의 기대값 (expected value of linear functions of random variables)****정리(THEOREM)**

$Y_1, Y_2, \dots, Y_n \sim (\mu_i, ?), X_1, X_2, \dots, X_m \sim (\xi_j, ?)$  라 하고  $U = \sum_{i=1}^n a_i X_i, V = \sum_{j=1}^m b_j Y_j$  라 놓자.

$$(1) E(U) = \sum_{i=1}^n a_i E(X_i) = \sum_{i=1}^n a_i \mu_i$$

$$(2) V(U) = \sum_{i=1}^n a_i^2 V(Y_i) + 2 \sum_{i < j} a_i a_j Cov(Y_i, Y_j)$$

$$(3) Cov(U, V) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(Y_i, X_j)$$

$$(이변량) V(X_1 \pm X_2) = V(X_1) + V(X_2) \pm 2COV(X_1, X_2)$$

**EXAMPLE 5.23**

결합확률밀도함수  $f(y_1, y_2) = 3y_1, 0 \leq y_2 \leq y_1 \leq 1$  일 경우  $(X_1 - X_2)$ 의 분산을 구하시오.

0.06

$$f(y_1) = 3y_1^2, 0 \leq y_1 \leq 1, \quad f(y_2) = (3/2)(1 - y_2^2), 0 \leq y_2 \leq 1$$

$$E(Y_1) = 3/4, E(Y_2) = 3/8, \quad E(Y_1^2) = 3/5, E(Y_2^2) = 1/5 \quad V(Y_1) = 0.04, V(Y_2) = 0.06, COV(Y_1, Y_2) = 0.02$$

**EXAMPLE 5.24**

확률표본  $X_1, X_2, \dots, X_n$ 은 서로 독립이고 평균은  $E(Y_i) = \mu_i$ , 분산은  $V(Y_i) = \sigma^2$ 을 갖는다. 표본 평균  $\bar{X} = \sum X_i / n$ 의 평균과 분산을 구하시오.

$$\mu, \sigma^2/n$$

**EXAMPLE 5.25**

확률변수  $Y$ 는 성공 확률이  $p$ 인 베르누이 시행을  $n$ 번 했을 때 성공의 회수이다. 표본

비율  $\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$ ,  $i$ -번째 실험의  $X_i = 0$ (실패),  $1$ (성공)의 평균과 분산을 구하시오.

$$p, pq/n$$

**EXAMPLE 5.26**

결합밀도함수  $f(x_1, x_2) = 4x_1x_2, 0 \leq x_1, x_2 \leq 1$ 이다.  $E(X_1 - X_2)$ ,  $V(X_1 - X_2)$ 을 구하시오.

$$0, 1/9$$

**EXAMPLE 5.27**

$E(X_1) = 2, E(X_2) = -1, E(X_3) = 4, V(X_1) = 4, V(X_2) = 6, V(X_3) = 8$ 이다.  
 $COV(X_1, X_2) = 1, COV(X_1, X_3) = -1, COV(X_2, X_3) = 0$   
 $E(3X_1 + 4X_2 - 6X_3)$ ,  $V(3X_1 + 4X_2 - 6X_3)$ 을 구하시오.

**HOMEWORK #14-3**

Let  $X_1, X_2$  be uncorrelated random variable. Find the covariance and correlation of  $W = X_1 + X_2, U = X_1 - X_2$  in terms of the variance and covariance of  $X_1, X_2$ .

**HOMEWORK #14-4**

Let  $f(y_1, y_2) = 2, 0 \leq y_1, y_2 \leq 1; 0 \leq y_1 + y_2 \leq 1$ . Find  $E(Y_1 + Y_2)$  and  $V(Y_1 + Y_2)$ .

**HOMEWORK #14-5**

Let  $f(y_1, y_2) = (1/8)y_1 e^{-(y_1+y_2)/2}, 0 \leq y_1, y_2$ . Find  $E(C)$  and  $V(C)$  where  $C = 50 + 2Y_1 + 4Y_2$ .

**5.8 조건부 기대치(Conditional Expectations)****정의(Definition)**

$$\begin{aligned} E(g(X_1) | X_2 = x_2) &= \int g(x_1) f_{1|2}(x_1 | x_2) dx_1 \text{ (연속형)} \\ &= \sum_{x_1} g(x_1) p_{1|2}(x_1 | x_2) \text{ (이산형)} \end{aligned}$$

$$V(X_1 | X_2 = x_2) = E(X_1^2 | X_2 = x_2) - E(X_1 | X_2 = x_2)^2$$

**정리(Theorem)**

- (1)  $E(X) = E[E(X | Y)]$
- (2)  $V(X) = E[V(X | Y)] + V[E(X | Y)]$

**Prove**

$$\begin{aligned} E(Y_1) &= \int y_1 f(y_1) dy_1 = \int y_1 \int f(y_1, y_2) dy_2 dy_1 \\ &= \int \int y_1 f(y_1 | y_2) f(y_2) dy_1 dy_2 \\ &= \int E(Y_1 | Y_2 = y_2) f(y_2) dy_2 \end{aligned}$$

$V(X)$ 에 대한 증명은 다음 사실을 이용하여 가능하다.

$$V(X | Y) = E(X^2 | Y) - E(X | Y)^2, \quad E(V(X | Y)) = E\{E(X^2 | Y)\} - E(E(X | Y))^2,$$

$$V(E(X | Y)) = E\{E(X | Y)^2\} - E(E(X | Y))^2 \quad (Q.E.D.)$$

**EXAMPLE 5.27**

결합확률밀도  $f(x_1, x_2) = 1/2, 0 \leq x_1 \leq x_2; 0 \leq x_2 \leq 2$ .  $E(Y_1 | Y_2 = 1.5)$ .

0.75

$$f(y_1 | y_2) = \frac{1}{y_2}, 0 \leq y_1 \leq y_2 \leq 2, \quad E(Y_1 | Y_2 = y_2) = \frac{y_2}{2}$$

**EXAMPLE 5.28**

확률변수  $Y$ 는 이항분포  $Binomial(n=3, p)$  따른다. 모집단 비율  $p$ 는 확률변수로 균일분포  $Uniform(0,1)$ 을 따른다고 한다.  $E(Y)$ ,  $V(Y)$ 을 구하시오.

**TIP** (1)  $E(X) = E[E(X | Y)]$  (2)  $V(X) = E[V(X | Y)] + V[E(X | Y)]$

1.5 / 1.25

**HOMEWORK #15-1**

확률변수  $Y$ 가 모수가  $\lambda$ 인 포아송 분포를 따른다.  $Y \sim Poisson(\lambda)$ . 모수  $\lambda$ 의 확률밀도함수는  $f(\lambda) = e^{-\lambda}, \lambda \geq 0$ 이다.  $Y$ 의 기대치와 분산을 구하시오.

**5.9 다항 분포 (Multinomial Probability Distribution)****다항 실험 Multinomial experiment**

① 실험은  $n$ 개의 동일하고(identical), 독립(independent)인 시행이다.

② 각 시행의 결과는  $k$ 개이고 각 결과의 성공 확률은  $p_i$ 이다.  $\sum_{i=1}^k p_i = 1$

③ 확률변수  $X_1, X_2, \dots, X_k$ ;  $X_i$  는 결과  $i$  발생 수이다.  $\sum_{i=1}^n X_i = n$ .

확률변수  $X_1, X_2, \dots, X_k$  의 결합확률밀도함수는

$$p(x_1, x_2, \dots, x_n) = \binom{n}{x_1 x_2 \dots x_n} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, \sum p_i = 1, \sum x_n = n$$

### 정리 (THEOREM)

$$(1) E(Y_i) = np_i, V(Y_i) = np_i q_i$$

$$(2) COV(Y_s, Y_t) = -np_s p_t, \text{ for } t \neq s$$

### Proof

(1)  $X_i \sim \text{Binomial}(n, p_i)$  는 쉽게 알 수 있다.

(2) 만약  $i$ -번 시행에서 결과  $s$  가 일어나면  $U_i = 1$ , 그렇지 않으면  $U_i = 0$  이라 정의하고

$X_s = \sum_{i=1}^n U_i$  이라 하자. 그리고  $j$ -번 시행에서 결과  $t$  가 일어나면  $W_j = 1$ , 그렇지 않으면

$W_j = 0$  이라 정의하고  $X_t = \sum_{j=1}^n W_j$  이라 하자.

$$\begin{aligned} COV(Y_s, Y_t) &= \sum_{i=1}^n \sum_{j=1}^n COV(U_i, W_j) \\ &= \sum_{i=1}^n COV(U_i, W_i) + \sum_{i \neq j} COV(U_i, W_j) \quad \boxed{Q.E.D.} \\ &= -np_s p_t \text{ why? } COV(U_i, W_i) = E(U_i W_i) - E(U_i)E(W_i) = 0 - p_s p_t \end{aligned}$$



### HOMEWORK #15-2

Let  $p(y_1, y_2) = \binom{n}{y_1 y_2 (n - y_1 - y_2)} p_1^{y_1} p_2^{y_2} (1 - p_1 - p_2)^{(n - y_1 - y_2)}$ . Find the marginal pdf of  $Y_1$ .



### HOMEWORK #15-3

A lot of items contains 10% with exactly one defect, 5% with more than one defect, and the remainder with no defect. Ten item are randomly selected from the lot. If  $Y_1$  denote the number of items with one defect, and  $Y_2$  the number of items with one more that one defect. Suppose the repair cost are  $Y_1 + 3Y_2$ . Fine the mean and variance of the repair cost. Let  $Y$  be the number of items among ten that have at least one defect. Find  $P(Y = 2)$  and  $P(Y \leq 1)$

### 5.10 결합 적률생성함수 (Joint MGF)

**Recall**  $M(t) = E(e^{tX})$ : (일변량) 확률변수  $Y$ 의 MGF,

- 적률생성함수는  $t$ 의 함수이다.
- 적률을 구하는데 사용한다.  $M^{(k)}(t=0) = E(X^k)$ .
- MGF의 유일성(uniqueness): 적률생성함수가 같은 확률변수는 동일한 분포함수를 갖는다. 이를 이용하여 확률변수 함수(function of random variable)가 어떤 분포를 갖는지 알게 된다.

#### 정의(DEFINITION)

확률변수  $X_1, X_2$ 의 결합 적률생성함수(joint MGF)는  $M_{X_1, X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2})$ 이다.

$X_1, X_2$  주변(marginal) MGF는 각각  $M_{X_1, X_2}(t_1, 0), M_{X_1, X_2}(0, t_2)$ 로 정의한다.

#### THEOREM①

상수  $a, b$ 일 경우  $Y = a + bX$ 의 적률생성함수는  $M_Y(t) = M_{a+bX}(t) = e^{at} M_X(bt)$ 이다.

**Proof**

$$M_Y(t) = M_{a+bX}(t) = E(e^{t[a+bX]}) = e^{at} E(e^{(tb)X})$$



#### EXAMPLE 5.29

$Z \sim \text{Normal}(0,1)$ 일 경우  $X = \mu + \sigma Z$ 는 어떤 분포를 따르는가?

**THEOREM②**

두 확률변수  $X_1, X_2$  일 경우  $M_{X_1, X_2}(t_1, t_2) = M_{X_1}(t_1)M_{X_2}(t_2)$  이다.

**Proof**

$$M_{X_1, X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2}) = E(e^{t_1 X_1})E(e^{t_2 X_2}) = M_{X_1}(t_1)M_{X_2}(t_2)$$

**THEOREM③**

확률표본  $X_1, X_2, \dots, X_n (\Leftrightarrow X_i \stackrel{\sim}{\sim} f(x), M_X(t))$ 에서  $U = \sum_{i=1}^n a_i X_i$ 의 적률생성함수는

$$M_U(t) = \prod_{i=1}^n M_X(a_i t).$$

**HOMEWORK #15-4**

Prove the above theorem③.

**EXAMPLE 5.30**

결합확률밀도함수  $f(y_1, y_2) = e^{-y_1}, 0 \leq y_2 \leq y_1$  이다.

(1)  $M_{Y_1, Y_2}(t_1, t_2) = \frac{1}{(1-t_1-t_2)(1-t_1)}$ 임을 보이시오.

(2) 확률변수  $Y_1, Y_2$ 의 주변 MGF 구하고 이를 이용하여 어떤 분포를 따르는지 밝히시오.

**EXAMPLE 5.31**

$X_1, X_2, \dots, X_n$ 는  $Normal(\mu, \sigma^2)$  으로부터 확률표본일 경우.

(1)  $\sum_{i=1}^n X_i$  는 어떤 분포함수를 따르는가?

(2)  $\bar{X} = \sum_{i=1}^n X_i / n$  는 어떤 분포함수를 따르는가?

**APPENDIX1**

$X \sim Exponential(\beta)$  의 적률생성함수를 구하시오.

$$\begin{aligned} m(t) = E(e^{tX}) &= \int_0^{\infty} e^{tx} \frac{1}{\beta} e^{-x/\beta} dx = \frac{1}{\beta} \int_0^{\infty} e^{-x(1-\beta t)/\beta} dx \quad (\text{Since } \beta/(1-\beta t) > 0 \Rightarrow t < 1/\beta) \\ &= \frac{(\frac{\beta}{1-\beta t})}{\beta} \int_0^{\infty} \frac{1}{(\frac{\beta}{1-\beta t})} e^{-x(1-\beta t)/\beta} dx = \frac{1}{1-\beta t}. \end{aligned}$$

Q.E.D.

**HOMEWORK #15-5**

$X_1, X_2, \dots, X_n$  은  $Exponential(\beta)$  일 경우  $U = \sum_{i=1}^n X_i$  는 어떤 분포를 따르는가?

**APPENDIX2**

Find the MGF of  $X \sim Normal(\mu, \sigma^2)$ .

우선  $(X - \mu)$ 의 적률생성함수를 구하자.

$$\begin{aligned} M_{X-\mu}(t) &= \int_{-\infty}^{\infty} e^{t(x-\mu)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{y^2}{2\sigma^2}} dy \\ &= e^{t^2\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\sigma^2 t)^2}{2\sigma^2}} dy = e^{t^2\sigma^2/2} \end{aligned}$$

$$M_{X-\mu}(t) = E(e^{t(X-\mu)}) = E(e^{tX})e^{-t\mu} = M_X(t)e^{-t\mu}. \text{ 그러므로 } M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}) \quad \boxed{Q.E.D}$$



### APPENDIX3

$X_1, X_2$ 의 이변량 정규확률밀도함수는

$$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{z}{2(1-\rho^2)}\right],$$

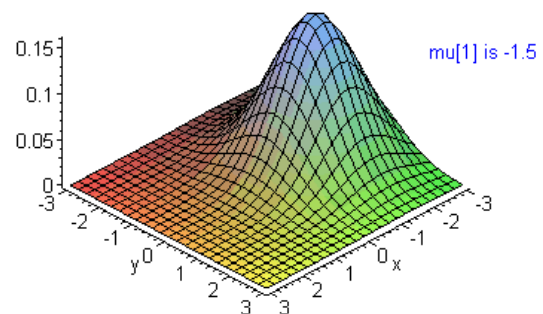
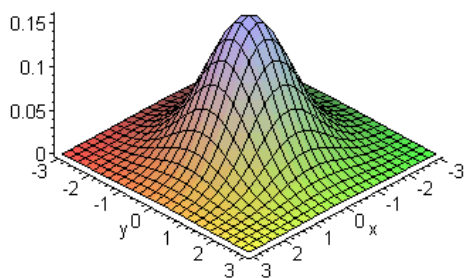
$$z \equiv \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}, \quad \rho \equiv \text{cor}(x_1, x_2) = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

- (1)  $X_i$ 의 주변확률밀도함수는  $N(\mu_i, \sigma_i^2)$ .
- (2)  $\rho$ 는 상관계수이다.
- (3) Box-Muller Transformation:  $Y_1, Y_2 \sim (iid) \text{Uniform}(0,1)$  일 경우

$X_1 = (-2 \ln Y_1)^{1/2} \cos(2\pi Y_2)$  은 이변량 정규분포를 따른다.  
 $X_2 = (-2 \ln Y_1)^{1/2} \sin(2\pi Y_2)$

Bivariate Normal

mu[1] is changing!



<http://www2.kenyon.edu/People/hartlaub/MellonProject/Bivariate2.html>